Ph.D. Final Synopsis

Title

Ancestral Origin and Genomic Diversity of Koṅkaṇī Speakers from Koṅkaṇ-Malabar region of India: A Molecular Approach

SUBMITTED FOR THE AWARD OF
DOCTOR OF PHILOSOPHY
DEGREE IN
APPLIED ZOOLOGY

Submitted by

JAISON JEEVAN SEQUEIRA

No. MU/EXB/Ph.D./CR.15/A.Z./2021-2022/E.25 dated 26.10.2021
Ph.D. Registration: 06.08.2019

Research Supervisor

Dr. MOHAMMED SHAFIUL MUSTAK. B
Professor
Department of Applied Zoology
Mangalore University

Place of Investigation

Department of Applied Zoology
Mangalore University
Mangalagangothri
Karnataka-574 199, India

AUGUST – 2023

1. ABSTRACT

Koṅkaṇī speakers of the Koṅkaṇ-Malabar region of India can be classified broadly under Sārasvata and non-Sārasvata groups. Gauḍa Sārasvata Brahmins (GSB), the Citrapur Sārasvata Brahmins (CSB) and Rājāpur Sārasvata Brahmins (RSB) who identify themselves as the descendants of the community that originated on the banks of river Sarasvati form the Sārasvata group. The non-Sārasvata group includes Khārvi, Kuḍubi, Gãvkār, Pāgi, Roman Catholics, Veḷip, Koṅkaṇ Marāṭhā, Daivajña Brahmins and many other communities including the Siddis and the Navāyats, each with a different ethnic history. The present work, based on uniparental genetic markers, aims to reconstruct the ancestral origin of the Koṅkaṇī population, and explores the genetic diversity in Koṅkaṇī language speaking groups. The **first chapter** of this thesis describes the Koṅkaṇī population, its language and its geographical extent. Literature review summarizes a chronology of human migration and the genetic tools used to reconstruct ancestry.

The **second chapter** deals with the methodology used to achieve the objectives of this thesis. Over 500 individuals from the Koṅkaṇ-Malabar region of India were analysed. We used single nucleotide polymorphisms (SNPs) from mtDNA and Y chromosome as markers to reconstruct maternal and paternal ancestral lineages, respectively. Our findings are based on the pairwise differences, multivariate analysis and Bayesian analysis.

In the **third chapter** we describe the maternal genetic landscape in the Koṅkaṇī population. Our findings are mainly based on the mtDNA D-loop sequence variation. We also used whole mitogenome data to estimate the time to the most common recent ancestor (TMRCA) for major haplogroups found in this population. Overall, the maternal ancestry of the Sārasvata group shows affinity towards the north-western region of the Indian subcontinent. The non-Sārasvata group on the other hand, has lower maternal diversity and higher proportion of ancient autochthonous haplogroups. Particularly, Khārvi population shows greater affinity with the Andamanese mitogenome. Kuḍubi shows lower pairwise differences with ancient Iberian and Mediterranean mitogenomes suggesting a Late Pleistocene movement of West-Eurasian haplogroups. The presence of both pre-LGM (Last Glacial Maximum) and post-LGM Indian-specific M and West-Eurasian R haplogroups in indigenous populations is a striking observation also found in Koraga, Urali Kuruman, Vedda, Malekudiya, Lakshadweep etc. Furthermore, these populations often find affinities with the extant Iranian genepool. Therefore, it is plausible that some of the mitogenomes in these unique populations including Kuḍubi belonged to a different wave of modern humans into the Indian subcontinent, perhaps around the Last Glacial Maximum or before but certainly not a part of the Bronze age movement or recent admixture.

The **fourth chapter** deals with the paternal ancestral origin and diversity of the Koṅkaṇī population. The impact of the more recent admixture, evidently higher in the Sārasvata group, coincides with the timeframe of Aryan migration. Recent studies relate this event with the spread of Indo-European languages and the R1a haplogroup. Interestingly, Citrapur Sārasvata Brahmins remained genetically uninfluenced by this admixture whereas Gauḍa Sārasvata Brahmins, Rājāpur Sārasvata Brahmins, Roman Catholics and the non-Sārasvata groups show

R1a haplogroup in varying proportions. Prevalence of Yamnaya-specific R1 subclade in Citrapur Sārasvata Brahmins suggests that the Sārasvata paternal ancestry predates the lineage of R1a carrying Aryans that formed a major fraction of the present day North Indian genepool. In addition, the presence of haplogroups J2-M172, K-M9 and L1-M76 in the Sārasvata lineage suggests that their possible migratory range extends between the Levant, Near East and the Bactria-Margiana Archaeological Complex (BMAC) where they might have settled along the banks of river Sarasvati or river Helmand before migrating southward. Multivariate analysis also shows intra-population variation suggesting early separation and recent admixture in each of the subpopulations. Interestingly, the non-Sārasvata group has R1a frequencies comparable with those in the Sārasvata moieties. However, higher proportion of Indian-specific H1a-M69 and R2-M124 clusters them away from the Sārasvata group. Exaggerated R1a frequency in the non-Sārasvata groups could be attributed to more than one R1a influx at different time intervals. A downstream analysis of R1a subclades can provide a geographical distribution pattern of older and recent Y SNP markers. In conclusion, the maternal and paternal ancestral pattern observed in both Sārasvata and non-Sārasvata moieties suggests at least two major admixture events separated by the Last Glacial Maximum.

To further understand the genetic implications of ancient stratification within the Koṅkaṇī speaking subgroups, association studies were conducted using Sārasvata samples due to their purportedly known migratory history and the presence of exogamous subgroups. We used *TAS2R38* gene SNPs in order to study the local and global pattern of PTC sensitivity with an aim to find ancestry specific association. The **fifth chapter** concludes that the PTC (phenylthiocarbamide) sensitivity genotype in Koṅkaṇī speaking Sārasvata Brahmins is slightly similar to that of West Eurasian populations. The findings also suggest ancestry specific selection in *TAS2R38* gene variations for taste sensitivity at global level.

In the **sixth chapter**, we study the genetic implications of an interesting aspect of the Koṅkaṇī Sārasvata Brahmins – the gotra. For this purpose, we used the mtDNA and Y chromosome haplogroup distribution pattern and investigated its association with gotras. This system of governed marital institution showed significant genetic consequences shaping the paternal ancestry in the Koṅkaṇī speaking Sārasvata Brahmins. These findings draw a new parallel to the earlier hypothesis of localized genetic variation due to cultural processes in Indian populations by describing gotra system as a genetic barrier.

In summary, this thesis provides an overview of the genetic diversity within the Koṅkaṇī speaking population of the Koṅkaṇ-Malabar region. It also concludes that the Last Glacial Maximum played a major role in shaping the genepool of south-western Indian populations including the Koṅkaṇī population.

2. OBJECTIVES

Based on the literature survey we observed that a genetic study is necessary for Koṅkaṇī population. The lack of genetic evidence for the different versions of their origin and migration is considered to be the research gap. To address this gap following objectives were proposed.

1. To reconstruct maternal ancestry using mitochondrial DNA variation.
2. To reconstruct paternal ancestry using Y Chromosome SNPs.
3. To understand the *TAS2R38* genotype and its association with taste perception and other factors in a Koṅkaṇī Population.
4. To study the genetic footprint of gotra system in Koṅkaṇī Sārasvata Brahmins.

3. METHODOLOGY

3.1. General work flow

Sample collection

516 blood samples were collected from volunteers as per the approval guidelines of Institutional Human Ethics Committee, Mangalore University (MU-IHEC-2020-3 dated 21.08.2020). 2-3 ml blood was collected from unrelated healthy individuals residing along the Koṅkaṇ- Malabar coast of India after obtaining their written consent. All the donors were personally informed, and their ethnic history was documented. Blood sample collection was carried out by a trained phlebotomist as per the IHEC guidelines. Genomic DNA was extracted using phenol chloroform method and quantified using Nano Drop spectrophotometer.

Samples were randomly collected at the sample sites with prior knowledge of the abundance of Koṅkaṇī speakers. In order to reduce sampling errors, detailed discussions were held with the volunteers to know their lineage at least up to 3-4 generations. Collected samples included individuals from both Sārasvata (Gauḍa Sārasvata Brahmins (GSB), the Citrapur Sārasvata Brahmins (CSB) and Rājāpur Sārasvata Brahmins (RSB)) and non-Sārasvata groups (Khārvi, Kuḍubi, Gw̃vkār, Roman Catholics).

Genomic DNA Extraction

Reagents:

- Reagent A (Lysis buffer 1): 10 ml (10 mM) Tris HCl (pH 8.0), 109.54 gm (320 mM) Sucrose, 5ml (5mM) MgCl2, 1% Triton X.
- Reagent B (Lysis buffer 2): 40 ml (400 mM) Tris HCl (pH 8.0), 12 ml (60 mM) EDTA, 15ml (150 mM) NaCl, 1% SDS added after autoclaving.
- Reagent C: 5 M Sodium per chlorate.
- Tris Saturated Phenol: Phenol, 8-Hydroxy Quinoline, 0.5 Tris HCl (pH 8.0), 0.1 Tris HCl (pH 8.0).
- Chloroform: Isoamyl alcohol (24:1): 24 ml of chloroform was added to 1 ml Isoamyl alcohol.
- T.E. buffer (100 ml): 10 mM Tris HCl (pH 7.5), 1 mM EDTA (pH 8.0).

- 20% SDS: 20 g of SDS dissolved in 80 ml DDW at 65 ºC, volume made up to 100 ml.
- 70% ethanol: 70 ml of absolute alcohol in 30 ml DDW.

Protocol:

1. To the blood sample (2 ml), 4 volumes (6 ml) of reagent A was added in a polypropylene tube. Mixed gently till the solution becomes clear.
2. Centrifuged at 2500 rpm for 5 minutes to obtain a pellet, free from RBC's. The supernatant containing lysed RBC's was discarded carefully.
3. The pellet was disturbed thoroughly and half the (1 ml) volume of (as that of the blood sample) reagent B was added. The constituents were mixed thoroughly and gently by inverting for 3-4 minutes till the solution became viscous.
4. Reagent C (0.25 ml) was added one by fourth volume of reagent B and mixed gently for 3-4 minutes.
5. Equal volume (as that of reagent B+C) of phenol chloroform was added and centrifuged at 2500-3000 rpm for 7-8 minutes to get a separation of 3 layers viz. aqueous layer, protein layer, and solvent layer.
6. The aqueous layer was transferred carefully into another centrifuge tube using a cut end micro tip without disturbing the protein layer.
7. Equal volumes of chloroform was added to the supernatant and mixed gently for a minute and centrifuged at 2500 rpm for 5 minutes.
8. The aqueous layer was transferred to a fresh tube.
9. Two volumes of chilled absolute alcohol was added and mixed gently to precipitate the DNA.
10. DNA lump was spooled out into a fresh Eppendorf tube
11. Extracted DNA was washed twice with 70% alcohol and gives a short spin to remove any traces of alcohol.
12. The pellet was dried properly and dissolved in 50 to 100 micro litre TE Buffer.
13. The tubes were incubated at 55 °C for 45 minutes to enhance the dissolution.
14. The DNA samples were stored at 4 °C until quantification.

Gel electrophoresis was performed to check the quality of DNA samples. 0.8 gm of agarose (0.8%) was dissolved in 100 ml 0.5X TAE buffer in conical flask and was boiled to dissolve agarose completely. 0.7 µl of ethidium bromide was added. Gel was cooled down to 60 °C, poured onto a gel tray and was allowed to set. 1 µl of the sample was loaded onto the wells and electrophoresis was carried out at a constant voltage of 80V. After 30 minutes of run the gel was observed under Gel Doc.

DNA Quantification using Nanodrop Spectrophotometer:

DNA quantification was done using Nano Drop Spectrophotometer. Only those samples with an absorbance ratio (260/280 nm) of ~1.8 were chosen for PCR amplification. 20 ng/ µl working stock of samples was prepared from the initial stock of extracted DNA using PCR water. Initial stocks were stored at -80 ºC and the working stocks were stored at 4-8 °C for immediate use.

Genotyping

Polymerase chain reaction (PCR) was performed with relevant primers (Table 12) in order to process the samples for targeted sequencing. PCR amplification was performed in ABI Veritii (Applied Bio Systems) Thermocycler.

PCR conditions were as follows: initial denaturation at 95°C for 5 mins, denaturation at 95°C for 1 min, annealing (temperature corresponding to the given set of primers), extension at 72°C for 5 min. PCR product was run on 2% agarose gel for 20 min at 100V. Samples were assessed for successful amplification.

Sanger sequencing was performed at CCMB, Hyderabad. Sanger sequencing results were obtained in *.abi format. Mutations were scored based on the variation observed in the electropherogram. Concatenated sequences were generated for each individual sample in the form of *.fasta files. These files were used for further analysis described in each section below.

3.2. Mitochondrial DNA analysis

For mtDNA analysis, 225 samples were sequenced using 15 F/R, 18 F/R, 22F/R, 23 F/R and 24 F/R primers covering the mtDNA control region. The sequences were aligned against the revised Cambridge Reference Sequence (rCRS) using SeqMan NGen®. Version 7.0. DNASTAR Madison, WI. and UniPro Ugene software in batches of 10 to reduce multiple alignment errors. The variations were listed and run into Haplogrep 2.0 for haplogroup assignment. A quality check was performed to minimise errors.

Statistical Analysis

Data from previous studies on the Koṅkaṇī population, other populations and ancient DNA data from AmtDB were collected and used for comparative analysis. Haplogroup frequency distribution was observed using Microsoft Excel. DnaSP6 was used to measure the diversity indices and pairwise Fst was calculated using Arlequin 3.5. The genomic affinity of the Koṅkaṇī population was checked against the populations belonging to different regions, social structures and linguistic families reported in literature. A Fst heatmap was plotted based on pairwise Fst differences. To understand the extent of haplotype sharing between subgroups, a median joining network was plotted using PopART and NETWORK. Principal Component Analysis (PCA) plot for intra-population and inter-population analysis was plotted, using PAST 4 software and R studio. Neighbour Joining Tree was constructed using the same software. Variance between the groups and populations within the groups was measured with AMOVA in Arlequin 3.5 software. Bayesian Skyline Plots (BSP) were generated using BEAST 1.10.

BEAST v2.7.4 was used to do Bayesian age estimation. The haplogroups were used to filter the samples. For calibration, ancient mitogenomes from AmtDB that had better coverage and C14 dates were employed. As an outgroup, the L2c2 mitogenome of a Moreno person (PaMOR16007) was employed. 96 complete sequences were used in BEAST runs. Runs were performed using the HKY substitution model. For the HVS1 and HVS2 regions, rigorous

molecular clocks with mutation rates of 1.292 and 0.369 mutations/site/million years, respectively, were used (Connell et al. 2022). Coding region partition was run considering a mutation rate of 0.03 mutations/site/million years. The following settings were used in Treeannotator to create a consensus tree: Common Ancestor Heights, 10% burnin. Figtree software was used to visualize and annotate the phylogenetic tree.

3.3. Y Chromosome SNP analysis

For paternal ancestry analysis, DNA from 273 males was amplified and sequenced for 19 biallelic Y SNP markers covering all the major haplogroups found in the Indian subcontinent using Sanger sequencing method. UniPro Ugene software was used to align the sequences against the reference sequences accessed from https://www.ensembl.org and variations were scored. Haplogroups were assigned as per ISOGG Haplotree and Y Chromosome STR Haplotype reference database (https://yhrd.org).

For downstream analysis of R1-M173 haplogroup observed in CSB, we performed YSTR analysis. AmpFlSTR® Yfiler® PCR Amplification Kit was used for amplification. Capillary electrophoresis was performed in Applied Biosystems® 3500/3500xL Genetic Analyzer. GeneMapper® ID Software was used to analyze the peaks and alleles were tabulated. Y haplogroups were predicted using Nevgen Y- Haplogroup Predictor online tool.

Statistical analysis

Haplogroups frequency distribution was calculated using Microsoft Excel data analysis package . Arlequin 3.5 was used to analyse pairwise Fst within population (Excoffier and Lischer, 2010). Interpopulation pairwise distance was plotted as heat map for better visualisation. Principal Component Analysis plot was plotted using R software and PAST4. Maps were generated using QGIS software.

3.4. PTC sensitivity analysis

For this analysis, 114 Sārasvata Brahmin samples with PTC taste data were employed. Gender, blood type, alcohol use, diabetes, lactase persistence, BMI, Rh factor, and other information were recorded during sample collection. A drop of PTC solution was placed on each participant's tongue tip to test their ability to taste PTC. There were established bitter tasters and non-tasters. After the test, participants were instructed to spit out their saliva.

TAS2R38 gene was amplified using PCR (Forward primer 5'-TAG GCAAAGAGCTGGATGCT-3' and Reverse primer 5'-ATTGGAAGGCTTTGTGAGGA-3'). SNPs determining PTC phenotypes (rs713598, rs1726866 and rs10246939) were scored and genotypes were tabulated.

Statistical analysis

Fisher's Exact test, the Odd Ratio test, and the Chi Square test were used to analyze the relationships between the variables. To determine the possibility of discovering PTC sensitivity in persons categorized based on the characteristics like gender, blood group, alcohol use,

diabetes, lactase persistence, BMI, Rh factor, etc., phenotype data was transformed to frequency distribution and an Odd Ratio analysis was performed. The genotype for the population and its subgroups was used to calculate the allele frequency of rs713598, rs1726866, and rs10246939. The SNP allele frequency was determined by manually counting the SNPs. The frequency distribution of haplotypes and diplotypes was tabulated. Chi Squared test and Fisher's Exact test were used to determine the relationship between PTC sensitive genotypes and various variables. P value was utilized for analysis. Where necessary, data was filtered to exclude factors with empty values. For analysis, only the two alleles that code for proline, alanine, valine, and isoleucine were taken into account (Sequeira *et al.*, 2023).

3.5. Gotra based analysis

We used 98 samples from males with gotra information for this analysis. Gotra details were provided by the participants in the questionnaire. Wherever Pravara information was provided, we derived the corresponding gotra using the classification system available in the literature. Even though gotra information from female participants was collected (father's gotra), female samples were not used for this analysis with a prior assumption of patrilineal lineage of gotras. However, mtDNA is present in males as well. Therefore, mtDNA haplogroup distribution pattern was considered for comparison.

Statistical analysis

In addition to the methods mentioned in 2.1 and 2.2, macrohaplogroup frequency distribution was calculated using Microsoft Excel. Diversity statistics were calculated using Arlequin Software. Principal Component Analysis plot and Neighbour Joining Tree for Inter Gotra analysis were plotted using PAST4 software. Chi Squared test and Fisher's Exact test were performed in R software.

4. RESULTS

Objective 1: To reconstruct maternal ancestry using mitochondrial DNA variation.

Background:

The human maternal genetic roots in India are spread up to 60000 years (Toomas Kivisild *et al.*, 1999) Although with limited power to reconstruct our distant past, mitochondrial DNA (mtDNA) could provide information of 100 to 200 thousand years of human evolution (Kivisild, 2015). mtDNA carries information of some of the earliest human migration events. However, reconstructing ancestry is a challenging task with each endogamous population carrying a piece of this information.

Earlier studies have broadly classified the Indian genepool into Ancestral North Indians, Ancestral South Indians, Ancestral Austro Asiatic and Ancestral Tibeto-Burman (Basu, Sarkar-Roy and Majumder, 2016). However, the maternal lineages in these pools are complex and interlinked. A common feature is the presence of M haplogroup in the Indian subcontinent with an earliest coalescence age of about 60,000 years (Macaulay *et al.*, 2005). For this reason, most M clades are considered Indian specific (Kumar *et al.*, 2008; Chandrasekar *et al.*, 2009). In

addition, we observe N, R and U lineages originating between 40-60,000 years ago (T Kivisild *et al.*, 1999; Palanichamy *et al.*, 2004). In the global context, maternal genepool can be broadly classified as but not limited to be West Eurasian, East Eurasian, Austronesian, Indian and African specific. In India most haplogroups belong to Indian specific, West Eurasian and East Eurasian clusters. Earlier studies on some of the south-western Indian populations have reported Indian specific and West Eurasian haplogroups (Thangaraj *et al.*, 2010; Palanichamy *et al.*, 2015; Sylvester *et al.*, 2018; Kumar *et al.*, 2022).

In this chapter we investigate the overall maternal lineage diversity of the Koṅkaṇī speaking population. Haplotype diversity, Nucleotide diversity and Tajima's D value are used to assess the impact of demographic changes in the past. In addition, the Bayesian Skyline Plots show the probable time of occurrence of such events. The next section describes the maternal genetic diversity in each of the studied subpopulations using the haplogroup distribution pattern. We also discuss about the clustering pattern between the subpopulations. In order to evaluate the affinity of the studied population with other Indian and global populations, we used two approaches. In the first approach, pairwise difference analysis was performed using the HVS1 sequence variation. In the second approach, haplogroup frequency distribution data was employed for Fst and multivariate analysis. Bayesian analysis of 96 whole mitogenomes provide an estimated time of coalescence for major haplogroups found in this population.

Results and discussion:

To understand the overall genetic diversity in this population we performed diversity statistics tests. Higher haplotype diversity (Hd) and lower nucleotide diversity (π) in the maternal lineage indicates small differences between haplotypes. It can also be attributed to rapid demographic expansion and accumulation of mutations after a population bottleneck (Searle 2000; Avise, Fisher, and Brick 2000; A. Sharma et al. 2020; W. S. Grant and Bowen 1997). This is evident through negative Tajima's D value and Fu's statistic (Tajima 1989; Y. X. Fu 1997) observed in this population. High prevalence of M haplogroup in this region is comparable with its ubiquitous presence throughout the Indian subcontinent. It is found in 60% Indians irrespective of castes and tribes they belong to. Koṅkaṇī population is no exception to it. Earlier studies suggest that this haplogroup could be indigenous to Indian subcontinent (Chandrasekar et al., 2009). However, it is still controversial whether the earliest dispersal for this haplogroup into the Indian subcontinent was out of Africa (Vincent et al. 2005; Kumarasamy et al. 2005; Mellars et al. 2013; Fernandes et al. 2012) or along a back migration from the southwest Asia into Africa (Marrero et al. 2016; T Kivisild 2003). R haplogroup is the second most abundant mitochondrial haplogroup (12.6%) found in Koṅkaṇī population. It is estimated to have been present in this subcontinent for about 65,000 years (Metspalu 2004). In the Indian context, R6, R7 and R8 subclades have a significant presence in the Austroasiatic tribes (Metspalu 2004; Chaubey et al. 2008; Kumarasamy Thangaraj et al. 2009; Sylvester et al. 2019). Interestingly, R8 haplogroup is found only in Khārvi and Kuḍubi population in the Koṅkaṇ region so far. It brings us closer to the assumption of historians who identify similarities between Mundari and Koṅkaṇī languages. A diffusion-based study of prehistoric population dynamics suggesting westward migration of eastern tribes and not from Goa (Vahia et al. 2017) and the presence of R8 in Malekudiya tribe in Karnataka (Sylvester et al. 2019) and other populations belonging to

Indo-European and Dravidian language families in low frequencies (Kumarasamy Thangaraj et al. 2009) substantiates that there were no traces of Ancient Austro Asiatic haplotypes in the Koṅkaṇ region at least until 40,000 YBP (approximate coalescent age of R8) and the movement of these early tribes carrying the R8 haplogroup extended southward through the Koṅkaṇ took place probably after the Last Glacial Maximum. This means that the Mundari link to Koṅkaṇī language is either incorrect or is a residue of this prehistoric migration or a back-migration from the Deccan region giving rise to a new language of the admixed lot. However, an assumption can still be made for a primordial tongue based on geographical factors that increase the probabilities for migration. One such factor is the presence of a plateau neighboring the Western Ghats. In our Y chromosome SNP analysis, we do not find any evidence for a Kuḍubi-Mundari link. Presence of U haplogroup at 12.3% in Kuḍubi population is comparable to caste populations rather than the tribes of India (Kivisild et al., 2003). Similarly, presence of J, H, HV and N haplogroups between 1-3% likens it to other caste populations. East Asian haplogroups D and G and West Asian haplogroup K is curiously present in negligible amounts in Koṅkaṇī population. This can be attributed to recent events of admixture.

To understand the stratification within the population we performed multivariate analysis based on the maternal haplotypes. We observed three distinct clusters. One of them all the Sārasvata groups along with the Roman Catholics. With a recent study involving autosomal data showing similar clustering and fairly well documented history, it is safe to assume that the recently converted Roman Catholics to a great extent belonged to the same lineage as the Gauḍa Sārasvata Brahmins (Kumar et al., 2021). The second cluster included the Kuḍubi and Gãvkār populations. The third cluster included the Khārvi tribe which stands out from the rest with highest frequency of an ancient haplogroup reported so far in India i.e. M57. From our Bayesian analysis it is clear that the M57a cluster found in Khārvi tribe branched out of other central Indian tribes like Nihal, Katkari, Kathakur and Dhongri Bhill. The presence of a ~25,000 year of R8 cluster in Khārvi and Kuḍubi groups suggest that this cluster belongs to the same source population that gave rise to the R8 cluster found in Urali Kuruman tribe during the Last Glacial Maximum. Our results are in contrast with the earlier hypothesis of autochthonous origin of R8 haplogroup (Thangaraj et al., 2009). The only other possibility of finding an R8 cluster outside the Indian subcontinent is through a back migration. However, our findings and the recent report of very high frequencies of R8 in Urali Kuruman tribe (Sylvester et al., 2019) suggest movement of both R8 and R30 haplogroups from the north west/near east region during or before the Last Glacial Maximum. Such a trend is explained by a delayed expansion hypothesis which assumes Middle East to be some kind of a hub for populations that migrated out of Africa (Xing et al., 2010; Vallini et al., 2022). Both intra and inter-population analysis suggest that the maternal ancestry of all Koṅkaṇī subgroups is distinct due to a peculiar pattern of M and R haplogroup subclades. In addition, we also find an early separation between the Sārasvata and non- Sārasvata clusters that became prominent after the introduction of social norms. However, the maternal lineage remained mostly uninfluenced by the recent demographic events which to a large extent was male mediated.

The Last Glacial Maximum has played an important role in shaping the maternal ancestry of many indigenous populations in the Indian subcontinent (Silva *et al.*, 2017). The same phenomenon is one of the events responsible for present day landscape of maternal lineages in Europe (Richards *et al.*, 2000; Jones *et al.*, 2015; Posth *et al.*, 2016; Fernández-López de Pablo *et al.*, 2019; Bortolini *et al.*, 2021). In the maternal lineage of the Koṅkaṇi populations, we find both pre and post LGM haplogroups. When individual subpopulations were studied, we observed unique distribution pattern in Sārasvata and Non Sārasvata groups. Interestingly, both these groups had pre and post-LGM haplogroups. Besides, we observed a few haplogroups that were present in both Sārasvata and Non Sārasvata groups. Based on the Bayesian analysis we could conclude that these populations arose from a diverse ancient source population, presumably in the Near East region, at the vicinity of the Last Glacial Maximum. When we performed a multivariate analysis on populations (from the Indian subcontinent) to check the pre and post LGM maternal haplogroup contribution, we observed a distinct scattering in the post LGM maternal ancestry. It is evident that the haplogroups like M3, M30, U7 and HV with TMRCA ~20,000 YBP in these populations are responsible for the scatter. In other words, the Last Glacial Maximum has influenced the distribution pattern of these haplogroups in the Indian subcontinent. Observation of increase in the effective population size around this time depth in the Bayesian skyline plots substantiates this hypothesis. In the context of Koṅkaṇī population, it appears that both Sārasvata and non-Sārasvata groups experienced a maternal lineage expansion in the post-LGM period. The clustering pattern observed in the subpopulations belongs to this time depth suggesting that the maternal genepool remained relatively uninfluenced by the recent admixture events.

Objective 2: To reconstruct paternal ancestry using Y Chromosome SNPs.

Background:

Koṅkaṇī speaking communities of the Koṅkaṇ-Malabar coast have different stories for their origin, migration and settlement in this region. Individuals of the Sārasvata group claims to be the descendent of Rig Vedic ancestors residing on the banks of river Sarasvati. They like to associate themselves with the lineage of Kashmiri Paṇḍits. Earlier studies involving Brahmin populations have reported higher frequency of R1a (Cordaux *et al.*, 2004; Sengupta *et al.*, 2006; Sharma, 2009; ArunKumar *et al.*, 2012). R1a haplogroups is found at moderate frequencies in certain tribes as well (ArunKumar *et al.*, 2012). This has led to a debate about the origin of this haplogroup which is linked with Aryan influx. Rival views (Sahoo *et al.*, 2006; Sengupta *et al.*, 2006; Sharma, 2009) disapprove this hypothesis of the association of movement of R1a with the spread of Indo-European languages and caste system. A seminal study of the bronze age Central and South Asia using ancient DNA concluded the spread of Steppe ancestry into South Asia likely carrying the Indo-European languages (Narasimhan *et al.*, 2019). They further associate this dispersal with Indus Valley decline resulting in the formation of Ancestral North Indian and Ancestral South Indian components in the present day Indian population (Reich *et al.*, 2009). Where does the Sārasvata group stand in this context? Y chromosome analysis will help us understand its position in this chain of past events.

The non-Sārasvata group on the other hand does not claim any sort of recent migration, except the Navayat and Siddi who are classified as Koṅkaṇī speaking communities of foreign origin. The common recent migration of both Sārasvata and non-Sārasvata groups is often linked with the Portuguese colonization of Goa. In this chapter we focus only on the ancient movements as we use Y SNP markers of basal clades. Based on the maternal lineage of the non- Sārasvata group, it is evident that although the Roman Catholics clustered closer to GSB, they experienced local admixture. We also observe a distinction between the maternal ancestry of Kuḍubi and Khārvi. Earlier studies (ArunKumar *et al.,* 2012; Chennakrishnaiah *et al.,* 2013) show relatively higher frequency of H1a, R1a and R2 in certain tribes. In addition, we also observe L1 and J haplogroups in South India reportedly associated with Neolithic agriculture. Khārvi is a fishing community whereas Kuḍubi found mostly in the interior parts of the Koṅkaṇ-Malabar coast is involved in agrarian activities. In the context of aforementioned debate, the ancestral affinities of these moieties will provide an overview of the demographic changes that impacted the paternal genepool of the South West coast of India.

Results and discussion:

Y Chromosome SNP analysis was performed to understand the paternal gene pool of Koṅkaṇī population. Frequency of haplogroup R was the highest at 59.8% followed by H (23.9%), L and J. Such dominance is apparent because the subgroups of Koṅkaṇī population did not include any tribal samples as most of them are identified today as caste groups. However, the higher frequency of H in Gãvkār (66.7%), Kuḍubi (57.4%) and Khārvi (37.9%) differentiates their paternal lineage from other subgroups. Haplogroup R1a seems to be distributed uniformly among all the subgroups with slightly higher occurrence in GSB and marginally lower in Khārvi, Kuḍubi and Gãvkār respectively. Our results show that the paternal gene pool of Koṅkaṇī population is a combination of both strictly Indian-specific haplogroup H (ASI component) and haplogroups R1a, R1, R2, L1 and J2 which could be attributed to demographic events placed at at-least two time-depths. The earliest migration as suggested by Narasimhan et al. (2019) brought the H haplogroup component into the Koṅkaṇ region which was followed by a post IVC decline migration that brought a more diverse ANI-ASI component which possibly paved way for the development of a language called Koṅkaṇī branching out from the Indo Aryan family of languages.

Interpopulation analysis results show a parallel between the age of paternal H1a, R1 and R1a haplogroups and maternal pre-LGM and post-LGM haplogroups. The age estimate for bifurcation of R1-M173 into R1a-M420 falls ~25,100 years ago (Underhill, 2015) and the origin of H1a-M69 and R2 lineages date back to the Late Pleistocene (Trivedi, 2008; Sharma *et al.,* 2012). If we consider LGM (~25000 years ago) as a time depth reference, then both Sārasvata and Non Sārasvata groups carry pre-LGM and post-LGM haplogroups. However, at micro level, each subgroup has unique pattern of distribution. Prevalence of R1-M173 in CSB is higher than RSB and GSB. Higher frequency of R1-M173 is found in Kashmiri Paṇḍits, Bihar Paswans, Gujrat Brahmins (Sharma, 2009) and Balochi population (Sengupta *et al.,* 2006). Similarly, R1a-M420 peaks in Sārasvata groups especially GSB which also carries H1a and pre-LGM maternal haplogroups like M57. Such older haplogroups are predominant in Kuḍubi and Khārvi subgroups. Major difference between the Sārasvata and Non-Sārasvata

groups is the presence of autochthonous R7, R5 and R8 maternal haplogroups in the latter. Traces of West Eurasian haplogroups are present in all the subgroups suggesting that all these groups diverged from an ancient population presumably in the North-Western or Near East region around the LGM. PCA places the Koṅkaṇī subgroups in between the ancient Western-Central-Eastern tribes and the North-western populations. In the spatial frequency distribution pattern we find that the paternal differences provide distinction to Sārasvata and Non-Sārasvata classification which exists in maternal lineage but marginally.

Objective 3: To understand the *TAS2R38* genotype and its association with taste perception and other factors in a Koṅkaṇī Population.

Background:

Variation in the TASR38 gene associated with the sensitivity towards phenylthiocarbamide (PTC) was first reported in 2003 (Kim *et al.*, 2003). Thereafter, single nucleotide polymorphisms (rs714598, rs1726866, rs10246939) in three amino acid positions at 49, 262 and 296, coding for Proline or Alanine, Alanine or Valine, and Valine or Isoleucine respectively, were discovered. The resulting haplotypes classified individuals into 'tasters' and 'non-tasters' based on their ability to sense the bitter tasting PTC compound (Kim and Drayna, 2005; Risso, Mezzavilla, *et al.*, 2016). The most common taster haplotype is PAV, whilst the most common non-taster haplotype is AVI. In addition, many rare haplotypes are found in various populations around the globe. Since the discovery of this genotype-phenotype association, numerous studies have been conducted to understand the physiological, dietary and disease-specific associations of the *TAS2R38* gene. Bitter taste perception is known to play a protective role against the ingestion of plant based toxic substances (Diószegi, Llanaj and Ádány, 2019). Along with haplotypes, diplotype combinations are also found to be associated with PTC sensitivity (Kim *et al.*, 2003).

In the global perspective, PAV tasters amongst Asians and Americans range in frequency between from ~64% to 68%, whereas amongst Europeans and Africans the frequency ranges between from ~45% to 50%. Non-taster AVI frequency peaks in Europeans (viz. ~49%), whilst in other populations the frequency varies between ~26% and 35% (Risso, Mezzavilla, *et al.*, 2016). Other haplotypes are found in compartively low frequencies, with AAI peaking in Africans (viz. ~13%). Earlier studies have suggested two hypotheses for the predominance of both the PAV and AVI haplotypes, i.e. balancing natural selection (Fisher, Ford and Huxley, 1939; Wooding *et al.*, 2004; Campbell *et al.*, 2012) and ancient balancing selection followed by recent demographic bottleneck events (Risso, Mezzavilla, *et al.*, 2016). Such a bimodal distribution is observed in PTC diplotypes as well. The AVI/AVI (non-taster), PAV/PAV (taster) and PAV/AVI or PAV/* heterozygote (taster) are the common compositions. Wooding et al. (2004) have proposed a fitness advantage of heterozygotes over homozygotes, linking this enhanced selective fitness with their higher global frequency (Wooding *et al.*, 2004).

In the last two decades, since the discovery of *TAS2R38* variation being associated with PTC sensitivity, very few studies have been conducted to explore the footprints of this gene in Indian populations. In an Asian Indian cohort (Pemberton *et al.*, 2008), the frequency of AVI was

found to be double the frequency of PAV. In the same study, more than half of the individuals were heterozygote tasters (AVI/PAV). Homozygote non-tasters (AVI/AVI) were almost four times more frequent than the homozygote tasters (PAV/PAV). Similar findings were reported in an Indian cohort (Deshaware and Singhal, 2017; Gupta *et al.*, 2018) as well as a tribal population on the west coast of India (Vinuthalakshmi, Sheik and Mustak, 2019). Overall, tasters are relatively less frequent in Indians (Pemberton *et al.*, 2008) compared to the global populations (Wooding *et al.*, 2004; Kim and Drayna, 2005) .

In the Indian context, PTC sensitivity is found to be associated with alcohol dependence and tobacco chewing (Vinuthalakshmi, Sheik and Mustak, 2019). Its association with body mass index (BMI) is disputed, as we find both significant (Gupta *et al.*, 2018) and insignificant (Deshaware and Singhal, 2017) results. Globally, association studies have turned up similar contradictions for alcohol dependency (Duffy *et al.*, 2004; Wang *et al.*, 2007). Smoking status is reported to be associated with *TAS2R38* in Hàn Chinese (Qi *et al.*, 2022) and North Americans of European ancestry, but not in North Americans of African ancestry (Risso, Kozlitina, *et al.*, 2016). Body mass index is associated with non-tasters in Italian females (Tepper *et al.*, 2008) and inversely associated in a Spanish cohort (Coltell *et al.*, 2019). Disease-specific association has also been reported for *TAS2R38*, including cancer (Sacerdote *et al.*, 2007; Carrai *et al.*, 2011; Choi *et al.*, 2016) and nonpolypoid chronic rhinosinusitis (Adappa *et al.*, 2016). Contradictions have also been reported (Timpson *et al.*, 2005; Gallo *et al.*, 2016; Lambert *et al.*, 2019). An association has also been observed with longevity (Melis *et al.*, 2019) and food preferences.

In this chapter, we try to understand the status of PTC sensitivity in Koṅkaṇī Sārasvata Brahmins and their subgroups. We also explore the association of PTC sensitivity with factors such as sex, blood group, alcohol consumption, diabetes, lactase persistence, BMI, Rh factor, etc. The Koṅkaṇī Sārasvata Brahmins constitute a group of Sārasvata Brahmins settled along the Koṅkaṇ Malabar coast of India. Koṅkaṇī Sārasvata Brahmins are a migrant population, who trace their origin to the archaeologically attested Sarasvatī river in the northwest of the subcontinent. Koṅkaṇī Sārasvata Brahmins mainly include the following genealogically based self-identifying subgroups or moeities: Gauḍa Sārasvata Brahmins (GSB), Citrapur Sārasvata Brahmins (CSB) and Rājāpur Sārasvata Brahmins (RSB). Koṅkaṇī Sārasvata Brahmins are an endogamous caste population, who speak the Koṅkaṇī language and represent a major section of the Indian Brahmin community.

Results and Discussion:

To date, this is the second study on a southwest Indian population reporting PTC sensitivity genotypes and associations. An earlier study on the Koragas, a tribal population of the Koṅkaṇ Malabar coast (Vinuthalakshmi, Sheik and Mustak, 2019) reported an association of PTC sensitivity with alcohol consumption and tobacco chewing. Similar observations were made in earlier studies (DiCarlo and Powers, 1998; Duffy *et al.*, 2004; Wang *et al.*, 2007). We did not find any such association with alcohol consumption. Lack of any association between PTC sensitivity and alcoholism has been reported earlier as well (Fischer *et al.*, 2014; Choi *et al.*, 2016). Also, our results correspond to an earlier study on an Indian cohort which did not report

any association with BMI or food preferences (Ooi *et al.*, 2010; Choi *et al.*, 2016; Deshaware and Singhal, 2017). Contradictions to our findings are reported in study (Gupta *et al.*, 2018). Besides a strong association of PTC sensitivity with taster and non-taster genotypes ($p<0.001$), we show an association between PTC diplotypes and the lactase persistence trait ($p<0.05$). The LCT gene associated with lactose intolerance and the *TAS2R38* gene have been studied together earlier for their involvement in food intake and a possible BMI association (Sacerdote *et al.*, 2007). The rs1446585 loci in the LCT gene and non-taster diplotypes have been shown to be associated with colorectal cancer (Carrai *et al.*, 2011). Both these genes are involved in BMI-related phenotypes (Corella *et al.*, 2011; Almon, Álvarez-León and Serra-Majem, 2012; Ortega *et al.*, 2016; Choi, 2019; Coltell *et al.*, 2019; Robino *et al.*, 2021). Therefore, our results can potentially be substantiated with a larger cohort and assessed for the genotype of this trait. The PROP sensitivity differences between males and females have been reported before (Bartoshuk, Duffy and Miller, 1994; Drewnowski, Kristal and Cohen, 2001; Fischer *et al.*, 2014). We report a trend towards association of the PTC sensitivity diplotype with sex. There are not enough data to conclude that PTC sensitivity is associated with blood groups. A few studies have correlated blood group (Malini *et al.*, 2010) and specifically blood group B (Leite *et al.*, 2018) with the PROP phenotype. Here, we report a trend towards association between blood groups and the PTC sensitivity diplotype.

Although a strong association was observed between the *TAS2R38* genotypes rs713598, rs1726866 and rs10246939, and tasters/non-taster phenotypes, the association profile for the rs713598 and rs1726866 genotypes needs to be studied in a larger cohort, as these markers deviated from the Hardy Weinberg equilibrium in the current study. A notable finding of our study is the significant association of rs10246939 and PTC sensitivity.

The frequencies of taster (57%) and non-taster (43%) phenotypes found in the study population are similar to those found in the earlier studies, which show a higher number of tasters as compared to non-tasters (Fareed *et al.*, 2012; Gupta *et al.*, 2018). However, the genotype frequency showed a higher number of non-taster haplotypes and diplotypes. Such an observation is made in earlier studies as well (Ghosh, 1973; Hakim *et al.*, 1973; Deshaware and Singhal, 2017; Vinuthalakshmi, Sheik and Mustak, 2019). The frequencies of taster haplotypes (PAV) and homozygote diplotypes (PAV/PAV) in Koṅkaṇī Sārasvata Brahmins are 8.2% and 8.3% respectively. These frequencies are lower than the global average, which is ~50% (Risso, Mezzavilla, *et al.*, 2016). Only 22.7% of the tasters had PAV/PAV, and only 13.3% carried PAV. We observed a 58.8% prevalence of the AVI haplotype, which is most prevalent in European populations (Risso, Mezzavilla, *et al.*, 2016). Interestingly, this non-taster haplotype peaked at 40% in the PTC tasters in our study population. Such contradictions are available in the literature (Kim *et al.*, 2003; Bufe *et al.*, 2005). We observe a higher frequency of single and double copies of AVV in Koṅkaṇī Sārasvata Brahmins. In tasters, the AVV haplotype is found at a frequency of 17.8%, and in non-tasters the AVV haplotype is found as a heterozygous diplotype (3.4% - 44.8%). The AVI/AVV combination is also prevalent in lactose intolerant individuals (27.3%). The AVI and AVV haplotypes and diplotypes are reported to be tasters, and the PAV and PAI haplotypes are reportedly non-tasters (Bufe *et al.*, 2005). We observed a significant association between haplotypes and PTC

sensitivity phenotype in Koṅkaṇī Sārasvata Brahmins. In our previous study, we identified the same correlation, i.e. the AVI/AVI haplotype in association with non-tasters, whilst the PAV/PAV haplotype was associated with tasters in the Koraga population (Vinuthalakshmi, Sheik and Mustak, 2019).

The minor allele frequency observed for rs713598, rs1726866 and rs10246939 in this population is comparable with the global datasets. The present generation of Koṅkaṇī Sārasvata Brahmins believe that their ancestors have migrated from the northern part of India. Genetic studies have supported this claim by finding Ancestral North Indian components in this population (Mascarenhas *et al.,* 2015; Kumar *et al.,* 2021). Moreover, Sārasvata Brahmins are found all over India with higher presence in the north west region. It is therefore a plausible hypothesis that Koṅkaṇī Sārasvata Brahmins who are presently found along the south west coast have migrated from the north. Earlier studies including ancient DNA studies have emphasised on Bronze age admixture of the Steppe pastoralists with the northwestern groups giving rise to Ancestral North Indians (Narasimhan *et al.,* 2019; van Driem, 2021). Coincidentally, the Sarasvatī river, which is often mentioned in the local folklore of the Koṅkaṇī Sārasvata Brahmins, is also believed to have dried up at the same time depth. Although the present study does not establish such a complex demographic event using SNP allele frequency variation, earlier studies have suggested ancestry specific selection for lactase persistence in specific ancestries (Gallego Romero *et al.,* 2012; Ranciaro *et al.,* 2014; Liebert *et al.,* 2017). In our study, both allelic and haplotypic variations observed in *TAS2R38* genotype of Koṅkaṇī Sārasvata Brahmins and other populations suggest ancestry specific selection in four broad clusters namely African, West Eurasian, Indian and East Asian. The alternate allele frequency for rs713598 (0.21), rs1726866 (0.44) rs10246939 (0.45) in Koṅkaṇī Sārasvata Brahmins is slightly less than the global average (0.422, 0.528 and 0.464 respectively) (Sayers *et al.,* 2022). Overall there appears to be a similarity in the allele frequencies between the study population and South Asian and West Eurasian datasets. When compared with the African, East Asian and West Eurasian datasets, the rs10246939 minor allele frequency in Koṅkaṇī Sārasvata Brahmins appeared more similar to that of West Eurasians. Although there exists a pattern in distribution, it does not provide enough support for the correlation of *TAS2R38* gene variation with micro-level demographic events, as we also observe a similar distribution pattern in South Asian populations with different ancestries. Nevertheless, these aspects can be explored further by genotyping larger cohorts of migrant populations in order to discern the genetic footprint of ancient migratory events on PTC sensitivity.

A famous example of a genotype-phenotype interaction is thought to be the association of the bitter taste receptor gene *TAS2R38* with PTC/PROP tasting ability. The relationship between *TAS2R38* gene variants and other factors like drinking, BMI, and malignancies has been studied by researchers. The global distribution of *TAS2R38* haplotypes exhibits a pattern that may have been altered by demographic events in populations' capacity for bitter taste. These gene differences could therefore be thought of as possible indicators of migration and adaption. The Koṅkaṇī Sārasvata Brahmin population offers an ideal case to study the genetic effects of migration because it is said to have a known history of migration. Their migration from the north to the south is sometimes linked with the time when the Indus Valley Civilization existed.

This notion has been amply supported by recent genomic analyses. Although the data we offer does not yet support the conventional wisdom, it also does not produce any evidence that is in opposition to it. With higher frequency of the non-taster AVI haplotype and similar minor allele frequencies for rs713598 and rs10246939, the Koṅkaṇī Sārasvata Brahmin group does indeed resemble the West Eurasian populations. We confirm the association of rs10246939 with PTC bitter taste sensitivity in an Indian population.

Objective 4: To study the genetic footprint of gotra system in Koṅkaṇī Sārasvata Brahmins.

Background:

The word gotra, which is etymologically related to the words *gau* (cow) and *trai* (protect), originally meant "cowshed" or "herd of cattle" in Sanskrit. Over time, the definition broadened to include the terms "subdivision," "tribe," "kin," and, more particularly, "exogamous lineage," which were employed by people of the Brahmin caste. As first described in the seminal studies of Shridhar Venkatesh Ketkar (Ketkar, 1909, 1911), the origin of the caste system in the wake of the migrations connected with the Indo-Aryan linguistic intrusion into the Indian Subcontinent unfolded as a complex process, which is now much better understood in ethnolinguistic terms (Mallory, 1989; van Driem, 2001, 2021). Ketkar and Karandikar (Karandikar, 1929) investigated the hypothesis that the patrilineal exogamic system of the Brahmans may have been adapted from pre-Aryan cultures. The gotra system of exogamous paternal lineages is historically linked to the emergence of Brahmanism, which occurred several centuries after the arrival of the Aryans in India (Ghurye, 1962; Madan, 1962). *Pravara* was originally a Sanskrit word that meaning "principal, chief, exalted" or "summons to perform rites." The meaning of this less common term expanded to include the connotation of "phratry" when it was employed in connection to the Saptari or seven legendary rishis (*ṛṣi*), who are thought to be the founders or *gotrakara* of the Brahmin lineages (Brough, 1947, 1953; Kosambi, 1950, 1953). Chentsalrao (1900) listed eighteen gotras that make up the patrilineal exogamic structure of the Brahmin caste, including Jamadagni, Bhāradvāja, Gautama, Kaśyapa, Vaśiṣṭha (also attested as: Vasiṣṭha), Agastya, Atri, Viśvāmitra, although the exact number of gotras varies depending on region, source, and tradition.

One must recreate the maternal and paternal ancestry of the current Brahmin community in order to comprehend the genetic consequences of such an ancient stratification. Our understanding of ancestral patterns on the Indian subcontinent has improved thanks to ground-breaking genome-wide studies (Moorjani *et al.*, 2013; Basu, Sarkar-Roy and Majumder, 2016; Narasimhan *et al.*, 2019). However, in order to comprehend the intricate patterns created by ancient stratification, micro-level investigations are required. The gotra system appears to have played a significant role in shaping ancestral patterning similar to endogamy that was introduced 70 generations ago probably at the start of the millenium (Basu, Sarkar-Roy and Majumder, 2016). In addition, there is ongoing debate over whether the diverse mosaic of communities on the Indian subcontinent share the same paternal ancestral patterns (Kumar, Banrida T. Langstieh, *et al.*, 2006).

The Sārasvata Brahmins, who may trace their ancestry to the banks of the river Sarasvati, are a prominent group of Brahmins in the context of the Koṅkaṇ coast. In the current study, we examine the genetic variation in Sārasvata Brahmins from the Koṅkaṇ region, who still steadfastly adhere to the gotra system.

Results and discussion:

The Brahmin communities in India are known to follow a strict set of restrictions to maintain exogamy. Three main restrictions include - marriage outside the caste or 'jaati', with the 'sapindas' and within the clan/gotras. Therefore, exogamous subgroups exist within each Brahmin community which are based on gotra system (Brough, 1953). Sārasvata Brahmin groups are spread along the western part of the Indian subcontinent. They claim to have originated along the banks of river Sarasvati. They also believe that their ancestors migrated southward due to the drying of the river and the drought that followed. One such group settled along the Koṅkaṇ coast namely, Koṅkaṇī Sārasvata Brahmins. In an endogamous migrant population like this, to understand the ways in which endogamy is maintained along with the rules of exogamy, we studied the molecular manifestation of the gotra system.

Due to the differential gotras distribution in the Koṅkaṇī Sārasvata Brahmins, all the subgroups clustered separately on the PCA scatterplot. CSB and NWSB clusters were due to the higher frequency of Vaśiṣṭha (83.3%) and Bhāradvāja (62.5%) gotras respectively. RSB had higher frequency of Bhāradvāja (27.3%) and Viśvāmitra (36.4%) while GSB showed higher frequency of Vaśiṣṭha (40%) and Jamadagni (27.1%). Similar distribution with higher frequencies of Bhāradvāja (Angiras group), Vaśiṣṭha and Jamadagni (Bhargava group) in Sārasvatas is reported earlier (Mascarenhas *et al.*, 2015). The Vaśiṣṭha gotra is reportedly linked with a unique cluster of western Brahmins namely "LPKSTR LPK" exhibiting Mitra–Varuna traditions followed by the early Iranic tribes (Mascarenhas *et al.*, 2015). Assimilation of specific gotras in subgroups is also reported in Bhargavas, Chaturvedis and Kanyakubjya and Sanadhaya brahmins. However, gotra based exogamy rules are increasingly violated (Bhatnagar and Agrawal, 2002).

In order to understand the genetic footprints of gotra system, we screened the maternal and paternal haplogroup distribution in Koṅkaṇī Sārasvata Brahmins using multivariate analysis. We observed a pattern of clustering in both maternal and paternal lineage, the latter being more distinct. Such correlation between the gotra and subgroup in paternal lineage is also observed in Cha Gham Patels (Pemberton *et al.*, 2012) and GSB samples from Lotli with Pai surnames (Mascarenhas *et al.*, 2015).

To understand the association between gotra, subgroup, maternal and paternal haplogroups we performed Chi Squared test. Strong association of gotras and paternal variation indicates that these exogamous units within Koṅkaṇī Sārasvata Brahmins indeed followed a strict marital pattern for generations that had genetic implications. In the Indian context, earlier studies were mostly based on historical records and surname classification (Bhatnagar and Agrawal, 2002; Devi *et al.*, 2008), with a few exceptions that included genetic analysis (Pemberton *et al.*, 2012; Mascarenhas *et al.*, 2015). Globally, various genetic studies on surname groups have been

conducted earlier showing influences on paternal lineage (King and Jobling, 2009), however, there is scarcity of such information for populations in the Indian subcontinent. We, for the first time, report that the gotra system shapes the paternal lineage in an Indian caste population i.e. Koṅkaṇī Sārasvata Brahmins. Comparison of clustering patterns based on haplogroup variation in gotras and subgroups showed similar results for paternal lineage. This finding is interesting because each of these subgroups have a different version of their historical origin. Our data, however, restricts us from substantiating an interlink between gotra and subgroup based on genetic variation. A STR analysis with a larger cohort of subgroups with gotra information is a future prospective.

We conclude that the gotra system has genetic implications shaping the paternal ancestry in the Koṅkaṇī Sārasvata Brahmins. Our findings draw a new parallel to the earlier hypothesis of localized genetic variation due to cultural processes in Indian populations (Kumar, Banrida T Langstieh, *et al.*, 2006) by describing gotra system as a genetic barrier. While the universality of gotra based genetic variation in the Indian subcontinent is yet to be determined, the Sārasvata Brahmins however represent a major proportion of Brahmin population that follow this type of social norm governing the marital institution. Hence, our findings are significant in hypothesising a universal paternal lineage pattern in caste populations.

5. SUMMARY

This thesis is focused on reconstructing the ancestral origin, and evaluating the genetic diversity of Koṅkaṇī speaking population residing on the south-western coast of India, precisely the Koṅkaṇ-Malabar region. We classified this diverse population broadly into Sārasvata and non-Sārasvata groups based on the existing versions of their origin. In order to fulfil the objectives of the present work, we sampled over 500 individuals from the Koṅkaṇ-Malabar region. Molecular and statistical tools employing mitochondrial DNA D-loop region sequence, Whole mitogenomes, Y chromosome SNPs and STRs, *TAS2R38* gene SNPS etc. were used to reconstruct ancestry, and evaluate ancestry specific variation. Some of the major findings of this thesis are as follows,

1. The maternal ancestry of the Sārasvata group (Gauḍa Sārasvata Brahmins (GSB), the Citrapur Sārasvata Brahmins (CSB) and Rājāpur Sārasvata Brahmins (RSB)) shows affinity towards the north-western region. The non-Sārasvata group (Khārvi, Kuḍubi, Gw̃vkār) has lower maternal diversity and higher proportion of ancient autochthonous haplogroups.

2. The Roman Catholics, although grouped under non-Sārasvata, along with local admixture shows similarity with the GSB. Among the non-Sārasvata moieties, Khārvi showed higher affinity towards the Andamanese component whilst Kuḍubi showed inclination towards the West-Eurasian component. We observe maternal haplotype sharing between GSB and Khārvi suggesting local admixture, possibly due to trade relations.

3. We report the highest M57 haplogroup frequency (57%) in Khārvi population.

4. The paternal ancestry of the Sārasvata and non-Sārasvata group differs greatly due to higher frequency of R1a in the former and H1a in the latter. The exaggerated R1a frequency in the non-Sārasvata groups can be attributed to more than one admixture event at different time intervals.

5. Yamnaya specific R1 subclade is predominant among the Sārasvata moieties, especially in CSB (61%) suggesting a pre-Bronze age origin near the Caucasus before intermixing with the R1a carriers in the north-west region.

6. *TAS2R38* genotype association with bitter taste sensitivity shows that the Sārasvata group has higher frequency of non-taster AVI haplotypes, a common feature in European populations. It further supports the existing hypothesis of their migration from the northern region.

7. We confirm the association of rs10246939 with PTC bitter taste sensitivity in Koṅkaṇī Sārasvata Brahmins.

8. The gotra system showed significant genetic implications shaping the paternal ancestry in the Koṅkaṇī speaking Sārasvata Brahmins.

In conclusion, we hypothesize that the Last Glacial Maximum had a greater impact on the genepool of south-western populations compared to the Neolithic or Bronze age dispersals. It appears that the recent admixture in the Koṅkaṇ-Malabar region was mostly male mediated and concerned with trade. The most recent dispersal of R1a possibly took place in the Bronze age carrying an Indo-Aryan language that mixed with the existing languages of the western region gradually forming the south-western branch including Koṅkaṇī.

6. REFERENCES

Adappa, N. D. *et al.* (2016) 'TAS2R38 genotype predicts surgical outcome in nonpolypoid chronic rhinosinusitis', *International Forum of Allergy & Rhinology*, 6(1), pp. 25–33. doi: https://doi.org/10.1002/alr.21666.

Almon, R., Álvarez-León, E. E. and Serra-Majem, L. (2012) 'Association of the European Lactase Persistence Variant (LCT-13910 C>T Polymorphism) with Obesity in the Canary Islands', *PLOS ONE*, 7(8), p. e43978. Available at: https://doi.org/10.1371/journal.pone.0043978.

Andrews, R. M. *et al.* (1999) 'Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA [5]', *Nature Genetics*, 23(2), p. 147. doi: 10.1038/13779.

ArunKumar, G. *et al.* (2012) 'Population Differentiation of Southern Indian Male Lineages Correlates with Agricultural Expansions Predating the Caste System', *PLOS ONE*, 7(11), p. e50269. Available at: https://doi.org/10.1371/journal.pone.0050269.

Baig, M. M., Khan, A. A. and Kulkarni, K. M. (2004) 'Mitochondrial DNA Diversity in Tribal and Caste Groups of Maharashtra (India) and its Implication on Their Genetic Origins', *Annals of Human Genetics*, 68(5), pp. 453–460. doi: https://doi.org/10.1046/j.1529-8817.2004.00108.x.

Bandelt, H. J., Forster, P. and Röhl, A. (1999) 'Median-joining networks for inferring

intraspecific phylogenies.', *Molecular Biology and Evolution*, 16(1), pp. 37–48. doi: 10.1093/oxfordjournals.molbev.a026036.

Barbieri, C. *et al.* (2012) 'Contrasting Maternal and Paternal Histories in the Linguistic Context of Burkina Faso', *Molecular Biology and Evolution*, 29(4), pp. 1213–1223. doi: 10.1093/molbev/msr291.

Barik, S. S. *et al.* (2008) 'Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands', *American Journal of Physical Anthropology*, 136(1), pp. 19–27. doi: https://doi.org/10.1002/ajpa.20773.

Bartoshuk, L. M., Duffy, V. B. and Miller, I. J. (1994) 'PTC/PROP tasting: Anatomy, psychophysics, and sex effects', *Physiology & Behavior*, 56(6), pp. 1165–1171. doi: https://doi.org/10.1016/0031-9384(94)90361-1.

Basu, A., Sarkar-Roy, N. and Majumder, P. P. (2016) 'Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure', *Proceedings of the National Academy of Sciences of the United States of America*, 113(6), pp. 1594–1599. doi: 10.1073/pnas.1513197113.

Bhatnagar, S. and Agrawal, S. (2002) 'Surname Endogamy among the Brahmin of India', *Current Sociology*, 50(6), pp. 853–861. doi: 10.1177/0011392102050006005.

Bortolini, E. *et al.* (2021) 'Early Alpine occupation backdates westward human migration in Late Glacial Europe', *Current Biology*, 31(11), pp. 2484-2493.e7. doi: 10.1016/j.cub.2021.03.078.

Bouckaert, R. *et al.* (2019) 'BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis', *PLOS Computational Biology*, 15(4), p. e1006650. Available at: https://doi.org/10.1371/journal.pcbi.1006650.

Brough, J. (1947) 'The early history of the gotras', *Journal of the Royal Asiatic Society*, 79 (1–2), pp. 76–90.

Brough, J. (1953) *The Early Brahmanical System of Gotra and Pravara: A Translation of the Gotra-Pravara-Manjari of Purusottama-Pandita*. Edited by J. Brough. Cambridge: Cambridge University Press.

Bufe, B. *et al.* (2005) 'The Molecular Basis of Individual Differences in Phenylthiocarbamide and Propylthiouracil Bitterness Perception', *Current Biology*, 15(4), pp. 322–327. doi: https://doi.org/10.1016/j.cub.2005.01.047.

Cabrera, V. M. *et al.* (2018) 'Carriers of mitochondrial DNA macrohaplogroup L3 basal lineages migrated back to Africa from Asia around 70,000 years ago', *BMC Evolutionary Biology*, 18(1), pp. 1–16. doi: 10.1186/s12862-018-1211-4.

Campbell, M. C. *et al.* (2012) 'Evolution of Functionally Diverse Alleles Associated with PTC Bitter Taste Sensitivity in Africa', *Molecular Biology and Evolution*, 29(4), pp. 1141–1153. doi: 10.1093/molbev/msr293.

Carrai, M. *et al.* (2011) 'Association Between TAS2R38 Gene Polymorphisms and Colorectal Cancer Risk: A Case-Control Study in Two Independent Populations of Caucasian Origin', *PLOS ONE*, 6(6), p. e20464. Available at: https://doi.org/10.1371/journal.pone.0020464.

Chandrasekar, A. *et al.* (2009) 'Updating Phylogeny of Mitochondrial DNA

Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor', *PLOS ONE*, 4(10), p. e7447. Available at: https://doi.org/10.1371/journal.pone.0007447.

Chennakrishnaiah, S. *et al.* (2013) 'Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India', *Gene*, 526(2), pp. 96–106. doi: 10.1016/j.gene.2013.04.074.

Chentsalrao, P. (1900) *Priniciples of Pravara and Gotra*. 2nd Editio. Mysore: Government Branch Press.

Choi, J.-H. *et al.* (2016) 'Genetic Variation in the TAS2R38 Bitter Taste Receptor and Gastric Cancer Risk in Koreans', *Scientific Reports*, 6(1), p. 26904. doi: 10.1038/srep26904.

Choi, J.-H. (2019) 'Variation in the TAS2R38 Bitterness Receptor Gene Was Associated with Food Consumption and Obesity Risk in Koreans', *Nutrients*. doi: 10.3390/nu11091973.

Coltell, O. *et al.* (2019) 'Association between taste perception and adiposity in overweight or obese older subjects with metabolic syndrome and identification of novel taste-related genes', *The American Journal of Clinical Nutrition*, 109(6), pp. 1709–1723. doi: 10.1093/ajcn/nqz038.

Cordaux, R. *et al.* (2004) 'Independent Origins of Indian Caste and Tribal Paternal Lineages', *Current Biology*, 14(3), pp. 231–235. doi: 10.1016/j.cub.2004.01.024.

Corella, D. *et al.* (2011) 'Association of the LCT-13910C>T Polymorphism With Obesity and Its Modulation by Dairy Products in a Mediterranean Population', *Obesity*, 19(8), pp. 1707–1714. doi: https://doi.org/10.1038/oby.2010.320.

Deshaware, S. and Singhal, R. (2017) 'Genetic variation in bitter taste receptor gene TAS2R38, PROP taster status and their association with body mass index and food preferences in Indian population', *Gene*, 627, pp. 363–368. doi: https://doi.org/10.1016/j.gene.2017.06.047.

Devi, N. K. *et al.* (2008) 'Surname Repetition in Three Population Groups of Andhra Pradesh: Kolam, Koya and Kuruma', *The Anthropologist*, 10(4), pp. 297–299. doi: 10.1080/09720073.2008.11891065.

DiCarlo, S. T. and Powers, A. S. (1998) 'Propylthiouracil Tasting as a Possible Genetic Association Marker for Two Types of Alcoholism 11Based on a dissertation submitted by S. T. D. to the faculty of St. John's University in partial fulfillment of the requirements for the PhD in psychology.', *Physiology & Behavior*, 64(2), pp. 147–152. doi: https://doi.org/10.1016/S0031-9384(98)00043-2.

Diószegi, J., Llanaj, E. and Ádány, R. (2019) 'Genetic Background of Taste Perception, Taste Preferences, and Its Nutritional Implications: A Systematic Review ', *Frontiers in Genetics* . Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01272.

Divisi, D. *et al.* (2017) 'Basic statistics with Microsoft Excel: a review', *Journal of thoracic disease*, 9(6), pp. 1734–1740. doi: 10.21037/jtd.2017.05.81.

Drewnowski, A., Kristal, A. and Cohen, J. (2001) 'Genetic Taste Responses to 6-n-Propylthiouracil Among Adults: a Screening Tool for Epidemiological Studies', *Chemical Senses*, 26(5), pp. 483–489. doi: 10.1093/chemse/26.5.483.

van Driem, G. L. (2021) *Ethnolinguistic Prehistory: The Peopling of the World from the Perspective of Language, Genes and Material Culture*. Leiden, The Netherlands: Brill. doi:

https://doi.org/10.1163/9789004448377.

van Driem, G. (2001) *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region : Containing an Introduction to the Symbiotic Theory of Language*. Brill (Handbook of Oriental Studies / 2: Indien). Available at: https://books.google.co.in/books?id=D7W6wAEACAAJ.

Duffy, V. B. *et al.* (2004) 'Bitter Receptor Gene (TAS2R38), 6-n-Propylthiouracil (PROP) Bitterness and Alcohol Intake', *Alcoholism: Clinical and Experimental Research*, 28(11), pp. 1629–1637. doi: https://doi.org/10.1097/01.ALC.0000145789.55183.D4.

Ehler, E. *et al.* (2019) 'AmtDB: a database of ancient human mitochondrial genomes', *Nucleic Acids Research*, 47(D1), pp. D29–D32. doi: 10.1093/nar/gky843.

Excoffier, L. and Lischer, H. E. L. (2010) 'Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology Resources*, 10(3), pp. 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.

Fairley, S. *et al.* (2020) 'The International Genome Sample Resource (IGSR) collection of open human genomic variation resources', *Nucleic Acids Research*, 48(D1), pp. D941–D947. doi: 10.1093/nar/gkz836.

Fareed, M. *et al.* (2012) 'Genetic study of phenylthiocarbamide (PTC) taste perception among six human populations of Jammu and Kashmir (India)', *Egyptian Journal of Medical Human Genetics*, 13(2), pp. 161–166. doi: https://doi.org/10.1016/j.ejmhg.2012.01.003.

Fernández-López de Pablo, J. *et al.* (2019) 'Palaeodemographic modelling supports a population bottleneck during the Pleistocene-Holocene transition in Iberia', *Nature Communications*, 10(1), p. 1872. doi: 10.1038/s41467-019-09833-3.

Fischer, M. E. *et al.* (2014) 'The Associations between 6-***n***-Propylthiouracil (PROP) Intensity and Taste Intensities Differ by ***TAS2R38*** Haplotype', *Lifestyle Genomics*, 7(3), pp. 143–152. doi: 10.1159/000371552.

Fisher, R. A., Ford, E. B. and Huxley, J. (1939) 'Taste-testing the Anthropoid Apes', *Nature*, 144(3652), p. 750. doi: 10.1038/144750a0.

Gallego Romero, I. *et al.* (2012) 'Herders of Indian and European Cattle Share Their Predominant Allele for Lactase Persistence', *Molecular Biology and Evolution*, 29(1), pp. 249–260. doi: 10.1093/molbev/msr190.

Gallo, S. *et al.* (2016) 'TAS2R38 taste receptor gene and chronic rhinosinusitis: new data from an Italian population', *BMC Medical Genetics*, 17(1), p. 54. doi: 10.1186/s12881-016-0321-3.

Gentula, M. C. and Nevski, A. (2015) *Y-DNA Haplogroup Predictor - NEVGEN*. Available at: https://nevgen.org/ (Accessed: 10 April 2023).

Ghosh, A. K. (1973) 'ABO Blood Groups and PTC Taste Sensitivity among the Kota of Nilgiri Hills', *Human Heredity*, 23(1), pp. 78–82. doi: 10.1159/000152557.

Ghurye, G. S. (1962) *Two Brahmanical Institutions: Gotra and Charana.* Bombay: Popular Prakashan.

Gupta, V. *et al.* (2018) 'Association of TAS2R38 polymorphism with measures of adiposity in Indian population', *Meta Gene*, 18, pp. 68–72. doi:

https://doi.org/10.1016/j.mgene.2018.08.001.

Hakim, S. M. A. *et al.* (1973) 'Blood Groups, Secretor Status and Ability to Taste Phenylthiocarbamide in some Muslim Groups', *Human Heredity*, 23(1), pp. 72–77. doi: 10.1159/000152556.

Hammer, Ø., Harper, D. A. T. and Ryan, P. D. R. (2001) 'PAST: Paleontological Statistics Software Package for Education and Data Analysis', *Palaeontologia Electronica*, 4(1), pp. 4–9.

Howe, K. L. *et al.* (2021) 'Ensembl 2021', *Nucleic Acids Research*, 49(D1), pp. D884–D891. doi: 10.1093/nar/gkaa942.

Jones, E. R. *et al.* (2015) 'Upper Palaeolithic genomes reveal deep roots of modern Eurasians', *Nature Communications*, 6(1), p. 8912. doi: 10.1038/ncomms9912.

Karandikar, S. V. (1929) *Hindu Exogamy*. Bombay: D.B. Taraporevala Sons & Co.

Ketkar, S. V. (1909) *The History of Caste in India: Evidence of the Laws of Manu on the Social Conditions in India during the Third Century A.D. Interpreted and Examined, with an Appendix on Radical Defects of Ethnology*. Ithaca, New York: Messrs. Taylor & Carpenter.

Ketkar, S. V. (1911) *An Essay on Hinduism, Its Formation and Future, Illustrating the Laws of Social Evolution as reflected in the History of the Formation of the Hindu Community*. London: Luzac and Co.

Kim, U. *et al.* (2003) 'Positional Cloning of the Human Quantitative Trait Locus Underlying Taste Sensitivity to Phenylthiocarbamide', *Science*, 299(5610), pp. 1221–1225. doi: 10.1126/science.1080190.

Kim, U. K. and Drayna, D. (2005) 'Genetics of individual differences in bitter taste perception: lessons from the PTC gene', *Clinical Genetics*, 67(4), pp. 275–280. doi: https://doi.org/10.1111/j.1399-0004.2004.00361.x.

King, T. E. and Jobling, M. A. (2009) 'What's in a name? Y chromosomes, surnames and the genetic genealogy revolution', *Trends in Genetics*, 25(8), pp. 351–360. doi: https://doi.org/10.1016/j.tig.2009.06.003.

Kivisild, T *et al.* (1999) 'Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages', *Current Biology*, 9(22), pp. 1331–1334. doi: https://doi.org/10.1016/S0960-9822(00)80057-3.

Kivisild, Toomas *et al.* (1999) 'The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World', *Genomic diversity: Applications in human population genetics*, pp. 135–152.

Kivisild, T. (2015) 'Maternal ancestry and population history from whole mitochondrial genomes', *Investigative Genetics*, 6(1), p. 3. doi: 10.1186/s13323-015-0022-2.

Kosambi, D. D. (1950) 'On the origin of Brahmin gotras', *Journal of the Bombay Branch of the Royal Asiatic Society*, XXVI: 21–8.

Kosambi, D. D. (1953) 'Brahmin clans', *Journal of the American Oriental Society*, 73 (4), pp. 202–208.

Kumar, L. *et al.* (2021) 'Dissecting the genetic history of the Roman Catholic populations of West Coast India', *Human Genetics*. doi: 10.1007/s00439-021-02346-4.

Kumar, L. *et al.* (2022) 'Genetic affinities and adaptation of the South West coast populations of India', *bioRxiv*, p. 2022.03.14.484270. doi: 10.1101/2022.03.14.484270.

Kumar, S. *et al.* (2008) 'The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage', *BMC Evolutionary Biology*, 8(1), p. 230. doi: 10.1186/1471-2148-8-230.

Kumar, V., Langstieh, Banrida T., *et al.* (2006) 'Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes', *PLoS Genetics*, 2(4), pp. 420–424. doi: 10.1371/journal.pgen.0020053.

Kumar, V., Langstieh, Banrida T, *et al.* (2006) 'Global Patterns in Human Mitochondrial DNA and Y-Chromosome Variation Caused by Spatial Instability of the Local Cultural Processes', *PLOS Genetics*, 2(4), p. e53. Available at: https://doi.org/10.1371/journal.pgen.0020053.

Lambert, J. D. *et al.* (2019) 'Bitter taste sensitivity, food intake, and risk of malignant cancer in the UK Women's Cohort Study', *European Journal of Nutrition*, 58(5), pp. 2111–2121. doi: 10.1007/s00394-018-1772-4.

Larruga, J. M. *et al.* (2017) 'Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area', *BMC Evolutionary Biology*, 17(1), pp. 1–15. doi: 10.1186/s12862-017-0964-5.

Leigh, J. W. and Bryant, D. (2015) 'POPART: Full-feature software for haplotype network construction', *Methods in Ecology and Evolution*, 6(9), pp. 1110–1116. doi: 10.1111/2041-210X.12410.

Leite, I. C. R. *et al.* (2018) 'Recognition of phenylthiocarbamide (PTC) in taste test is related to blood group B phenotype, females, and risk of developing food allergy: a cross-sectional Brazilian-based study', *Nutrition Research*, 52, pp. 22–38. doi: https://doi.org/10.1016/j.nutres.2017.12.013.

Liebert, A. *et al.* (2017) 'World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection.', *Human genetics*, 136(11–12), pp. 1445–1453. doi: 10.1007/s00439-017-1847-y.

Maca-Meyer, N. *et al.* (2003) 'Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography', *BMC Genetics*, 4(1), p. 15. doi: 10.1186/1471-2156-4-15.

Macaulay, V. *et al.* (2005) 'Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes', *Science*, 308(5724), pp. 1034–1036. doi: 10.1126/science.1109792.

Madan, T. N. (1962) 'Is the Brahmanic Gotra a Grouping of Kin?', *Southwestern Journal of Anthropology*, 18(1), pp. 59–77. doi: 10.1086/soutjanth.18.1.3629124.

Malini, S. S. *et al.* (2010) 'Expression of Genetic Traits among PTC Tasters and Non Tasters in Mysore Population, South India', *CAFET-INNOVA Technical Society*, p. 216.

Mallory, J. P. (1989) *In search of the Indo-Europeans*. London: Thames and Hudson.

Mascarenhas, D. D. *et al.* (2015) 'Genetic and Cultural Reconstruction of the Migration of an Ancient Lineage', *BioMed Research International*, 2015. doi: 10.1155/2015/651415.

Melis, M. *et al.* (2019) 'TAS2R38 bitter taste receptor and attainment of exceptional longevity', *Scientific Reports*, 9(1), p. 18047. doi: 10.1038/s41598-019-54604-1.

Metspalu, M. (2004) 'Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans', *BMC Genet.*, 5. doi: 10.1186/1471-2156-5-26.

Moorjani, P. *et al.* (2013) 'Genetic evidence for recent population mixture in India', *American journal of human genetics*. 2013/08/08, 93(3), pp. 422–438. doi: 10.1016/j.ajhg.2013.07.006.

Mustak, M. S. *et al.* (2019) 'The peopling of Lakshadweep Archipelago', *Scientific Reports*, 9(1), pp. 1–7. doi: 10.1038/s41598-019-43384-3.

Narasimhan, V. M. *et al.* (2019) 'The formation of human populations in South and Central Asia', *Science*, 365(6457), p. eaat7487. doi: 10.1126/science.aat7487.

Notes, E., Scheduled, O. N. and Tribes, S. (1961) 'CENSUS OF INDIA ETHNOGRAPHIC NOTES ON SCHEDULED CASTES ir', XI.

Okonechnikov, K. *et al.* (2012) 'Unipro UGENE: A unified bioinformatics toolkit', *Bioinformatics*, 28(8), pp. 1166–1167. doi: 10.1093/bioinformatics/bts091.

Ooi, S.-X. *et al.* (2010) 'Bitter Receptor Gene (TAS2R38) P49A Genotypes and Their Associations with Aversion to Vegetables and Sweet/Fat Foods in Malaysian Subjects', *Asia Pacific Journal of Clinical Nutrition*, 19(4), pp. 491–498. Available at: https://search.informit.org/doi/10.3316/informit.631770505094550.

Ortega, F. J. *et al.* (2016) 'Genetic variations of the bitter taste receptor TAS2R38 are associated with obesity and impact on single immune traits', *Molecular Nutrition & Food Research*, 60(7), pp. 1673–1683. doi: https://doi.org/10.1002/mnfr.201500804.

Palanichamy, M. G. *et al.* (2004) 'Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia.', *American journal of human genetics*, 75(6), pp. 966–978. doi: 10.1086/425871.

Palanichamy, M. G. *et al.* (2015) 'West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system', *Human Genetics*, 134(6), pp. 637–647. doi: 10.1007/s00439-015-1547-4.

Pemberton, T. J. *et al.* (2008) 'Prevalence of common disease-associated variants in Asian Indians', *BMC Genetics*, 9(1), p. 13. doi: 10.1186/1471-2156-9-13.

Pemberton, T. J. *et al.* (2012) 'Impact of restricted marital practices on genetic variation in an endogamous Gujarati group', *American Journal of Physical Anthropology*, 149(1), pp. 92–103. doi: https://doi.org/10.1002/ajpa.22101.

Pijpe, J. (2013) 'Indian Ocean crossroads: human genetic origin and population structure in the Maldives', *Am J Phys Anthropol*, 151. doi: 10.1002/ajpa.22256.

Posth, C. *et al.* (2016) 'Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe', *Current Biology*, 26(6), pp. 827–833. doi: 10.1016/j.cub.2016.01.037.

Qi, F.-Y. *et al.* (2022) 'Genetic variations in the bitter taste receptor gene TAS2R38 are related to cigarette smoking behavior in Han Chinese smokers', *Genes & Genomics*, 44(11),

pp. 1363–1374. doi: 10.1007/s13258-022-01311-3.

R Core Team (2021) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.

R Core Team (2022) ‘R: A Language and Environment for Statistical Computing’. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.

Ranasinghe, R. *et al.* (2015) ‘A study of genetic polymorphisms in mitochondrial DNA hypervariable regions I and II of the five major ethnic groups and Vedda population in Sri Lanka’, *Legal Medicine*, 17(6), pp. 539–546. doi: https://doi.org/10.1016/j.legalmed.2015.05.007.

Ranaweera, L. *et al.* (2014) ‘Mitochondrial DNA history of Sri Lankan ethnic people: Their relations within the island and with the Indian subcontinental populations’, *Journal of Human Genetics*, 59(1), pp. 28–36. doi: 10.1038/jhg.2013.112.

Ranciaro, A. *et al.* (2014) ‘Genetic origins of lactase persistence and the spread of pastoralism in Africa.’, *American journal of human genetics*, 94(4), pp. 496–510. doi: 10.1016/j.ajhg.2014.02.009.

Reich, D. *et al.* (2009) ‘Reconstructing Indian population history’, *Nature*, 461(7263), pp. 489–494. doi: 10.1038/nature08365.

Richards, M. *et al.* (2000) ‘Tracing European Founder Lineages in the Near Eastern mtDNA Pool’, *The American Journal of Human Genetics*, 67(5), pp. 1251–1276. doi: 10.1016/S0002-9297(07)62954-1.

Rieder, M. J. *et al.* (1998) ‘Automating the identification of DNA variations using quality-based fluorescence re-sequencing: Analysis of the human mitochondrial genome’, *Nucleic Acids Research*, 26(4), pp. 967–973. doi: 10.1093/nar/26.4.967.

Risso, D. S., Kozlitina, J., *et al.* (2016) ‘Genetic Variation in the TAS2R38 Bitter Taste Receptor and Smoking Behaviors’, *PLOS ONE*, 11(10), p. e0164157. Available at: https://doi.org/10.1371/journal.pone.0164157.

Risso, D. S., Mezzavilla, M., *et al.* (2016) ‘Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal’, *Scientific Reports*, 6(1), p. 25506. doi: 10.1038/srep25506.

Robino, A. *et al.* (2021) ‘Taste perception and expression in stomach of bitter taste receptor tas2r38 in obese and lean subjects’, *Appetite*, 166, p. 105595. doi: https://doi.org/10.1016/j.appet.2021.105595.

Rozas, J. *et al.* (2017) ‘DnaSP 6: DNA sequence polymorphism analysis of large data sets’, *Molecular Biology and Evolution*, 34(12), pp. 3299–3302. doi: 10.1093/molbev/msx248.

Sacerdote, C. *et al.* (2007) ‘Lactase Persistence and Bitter Taste Response: Instrumental Variables and Mendelian Randomization in Epidemiologic Studies of Dietary Factors and Cancer Risk’, *American Journal of Epidemiology*, 166(5), pp. 576–581. doi: 10.1093/aje/kwm113.

Sahoo, S. *et al.* (2006) ‘A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios’, *Proceedings of the National Academy of Sciences of the United States of America*. 2006/01/13, 103(4), pp. 843–848. doi: 10.1073/pnas.0507714103.

Sayers, E. W. *et al.* (2022) 'Database resources of the national center for biotechnology information.', *Nucleic acids research*, 50(D1), pp. D20–D26. doi: 10.1093/nar/gkab1112.

Sengupta, S. *et al.* (2006) 'Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists', *American Journal of Human Genetics*, 78(2), pp. 202–221. doi: 10.1086/499411.

Sequeira, J. J. *et al.* (2023) 'TAS2R38 bitter taste perception in the Koṅkaṇī Sārasvata Brahmin population', *Genes and Genomics*, (0123456789). doi: 10.1007/s13258-023-01409-2.

Shah, A. M. *et al.* (2011) 'Indian Siddis: African Descendants with Indian Admixture', *The American Journal of Human Genetics*, 89(1), pp. 154–161. doi: https://doi.org/10.1016/j.ajhg.2011.05.030.

Sharma, G. *et al.* (2012) 'Genetic Affinities of the Central Indian Tribal Populations', *PLOS ONE*, 7(2), p. e32546. Available at: https://doi.org/10.1371/journal.pone.0032546.

Sharma, I. *et al.* (2018) 'Ancient Human Migrations to and through Jammu Kashmir- India were not of Males Exclusively', *Scientific Reports*, 8(1), pp. 1–9. doi: 10.1038/s41598-017-18893-8.

Sharma, S. *et al.* (2005) 'Human mtDNA hypervariable regions, HVR I and II, hint at deep common maternal founder and subsequent maternal gene flow in Indian population groups', *Journal of Human Genetics*, 50(10), pp. 497–506. doi: 10.1007/s10038-005-0284-2.

Sharma, S. (2009) 'The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system', *J Hum Genet*, 54. doi: 10.1038/jhg.2008.2.

Silva, M. *et al.* (2017) 'A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals', *BMC evolutionary biology*, 17(1), p. 88. doi: 10.1186/s12862-017-0936-9.

Singh, G. *et al.* (2021) 'Genetic characterisation of the North-West Indian populations: analysis of mitochondrial DNA control region variations', *Annals of Human Biology*, 48(2), pp. 166–172. doi: 10.1080/03014460.2021.1879933.

Sylvester, C. *et al.* (2018) 'Neolithic phylogenetic continuity inferred from complete mitochondrial DNA sequences in a tribal population of Southern India', *Genetica*, 146(4–5), pp. 383–389. doi: 10.1007/s10709-018-0030-2.

Tepper, B. J. *et al.* (2008) 'Variation in the Bitter-taste Receptor Gene TAS2R38, and Adiposity in a Genetically Isolated Population in Southern Italy', *Obesity*, 16(10), pp. 2289–2295. doi: https://doi.org/10.1038/oby.2008.357.

Thangaraj, K. (2005) 'Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations', *Hum Genet*, 116. doi: 10.1007/s00439-005-1271-6.

Thangaraj, K. *et al.* (2005) 'Reconstructing the Origin of Andaman Islanders', *Science*, 308(5724), p. 996. doi: 10.1126/science.1109987.

Thangaraj, K. *et al.* (2010) 'The influence of natural barriers in shaping the genetic structure of Maharashtra populations', *PLoS ONE*, 5(12), pp. 1–9. doi: 10.1371/journal.pone.0015283.

Timpson, N. J. *et al.* (2005) 'TAS2R38 (phenylthiocarbamide) haplotypes, coronary heart disease traits, and eating behavior in the British Women's Heart and Health Study', *The American Journal of Clinical Nutrition*, 81(5), pp. 1005–1011. doi: 10.1093/ajcn/81.5.1005.

Trivedi, R. (2008) 'Genetic imprints of pleistocene origin of indian populations: a comprehensive Phylogeographic sketch of Indian Y-chromosomes', *Int J Hum Genet*, 8. doi: 10.1080/09723757.2008.11886023.

Underhill, P. A. (2015) 'The phylogenetic and geographic structure of Y-chromosome haplogroup R1a', *Eur J Hum Genet*, 23. doi: 10.1038/ejhg.2014.50.

Vinuthalakshmi, K., Sheik, N. and Mustak, M. S. (2019) 'TAS2R38 gene polymorphism and its association with taste perception, alcoholism and tobacco chewing among the Koraga -a primitive tribal population of Southwest coast of India', *Meta Gene*, 20, p. 100549. doi: https://doi.org/10.1016/j.mgene.2019.100549.

Wang, J. C. *et al.* (2007) 'Functional Variants in TAS2R38 and TAS2R16 Influence Alcohol Consumption in High-Risk Families of African-American Origin', *Alcoholism: Clinical and Experimental Research*, 31(2), pp. 209–215. doi: https://doi.org/10.1111/j.1530-0277.2006.00297.x.

Weissensteiner, H. *et al.* (2016) 'HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing', *Nucleic acids research*, 44(W1), pp. W58–W63. doi: 10.1093/nar/gkw233.

Wooding, S. *et al.* (2004) 'Natural Selection and Molecular Evolution in PTC, a Bitter-Taste Receptor Gene', *The American Journal of Human Genetics*, 74(4), pp. 637–646. doi: https://doi.org/10.1086/383092.

7. PUBLICATIONS

Related to the doctoral work

- **Sequeira, J. J**., Nizamuddin, S., van Driem, G., & Mustak, M. S. (2023). TAS2R38 bitter taste perception in the Koṅkaṇī Sārasvata Brahmin population. Genes & Genomics, 1-14. https://doi.org/10.1007/s13258-023-01409-2 (Published on 19 June 2023)
- Abstracts from the 55th European Society of Human Genetics (ESHG) Conference: e-Posters. Eur J Hum Genet 31 (Suppl 1), 91–344 (2023). https://doi.org/10.1038/s41431-023-01339-3 (EP20.001 Koṅkaṇ population along the southern route of earliest human migration)
- **Jaison J Sequeira** and Mohammed S Mustak (2022) **Genetic variation in Koṅkaṇī Sārasvata Brahmins due to gotra system** (in communication).
- **Jaison J Sequeira**, Lomous Kumar, George van Driem, Kumarasamy Thangaraj and Mohammed S Mustak (2022) **Mitochondrial DNA diversity in the Koṅkaṇī population on the west coast of India** manuscript is in submission stage.

Others

- Kumar, L., Chowdhari, A., Sequeira, J. J., Mustak, M. S., Banerjee, M., & Thangaraj, K. (2022). Genetic affinities and adaptation of the South West coast populations of India. bioRxiv, 2022-03. (in review)
- Singh, P. P., Rai, S. K., Chaubey, G., & BHU, S. C. (2023). Estimation of real COVID-19 cases in India during the first wave. IJID regions, 6, 80-83.
- Singh, P. P., Srivastava, A. K., Upadhyay, S. K., Singh, A., Upadhyay, S., Kumar, P., & BHU, S. C. (2021). The association of ABO blood group with the asymptomatic COVID-19 cases in India. Transfusion and Apheresis Science, 60(6), 103224.
- Alnoman Mundher Tayyeh, Jaison J Sequeira, Lomous Kumar, Idrees Babu, George van Driem and Mohammed Shafiul Mustak (2023) The Maternal ancestry of the Kavaratti Islanders and the Last Glacial Maximum aftermath (in review).

8. CONFERENCES

- Presented a paper (Online) on **"Koṅkaṇ population along the southern route of earliest human migration"** in the session "EP20 Population Genetics and Evolutionary Genetics" at the European Human Genetics Conference 2022 held at Vienna, Austria from June 11-14, 2022.
- Presented a paper (Online) on **"Genomic diversity in Koṅkaṇī population"** in the Technical Session: Physical Anthropology-I at the Indian Anthropology Congress 2022 held at Central University Hyderabad from February 21-23, 2022.
- Presented a poster titled "**The relevance of L-M20 haplogroup in the spread of Neolithic agriculture in the Dravidian heartland**" at the ADNAT conference held at BHU, Varanasi from March 10-12, 2023.
- Delivered an invited talk on **"The genetic landscape in the Koṅkaṇī population of Karnataka"** at the 46th Annual Conference of the Indian Society of Human Genetics (ISHG) held online from April 8-9, 2022.
- Delivered a talk on **"Koṅkaṇī people of Karnataka: A study of their origin, migration and settlement patterns"** at the Vishwa Koṅkaṇī Samaroh 2022 that took place from March 19-20, 2022 at Vishwa Koṅkaṇī Kendra, Mangaluru.